Semi-supervised Change Detection via Gaussian Processes

Keming Chen[1], Chunlei Huo[1], Zhixin Zhou[12], Hanqing Lu[1], Jian Cheng[1]

[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100190;

[2]Beijing Institute of Remote Sensing, Beijing, China, 100854

E-mail: {kmchen, clhuo, luhq, jcheng@nlpr.ia.ac.cn}, zhixin.zhou@ia.ac.cn

ABSTRACT

Change detection is one of the most important applications of the remote sensing technology. Compared with the supervised change detection approaches and the unsupervised ones, semi-supervised change detection methods which are more feasible at the operating level have received a great attention from the remote sensing community in recent years [1, 2]. The semi-supervised change detection method is based on a more realistic assumption that ground-truth information is available for at least a portion of the detected images. It involves a learning process which can take profit from the abundant cheep unlabeled data. Recently, kernel-based methods, such as inductive support vector machines (ISVMs) and transductive support vector machines (TSVMs), have been applied promisingly to remote sensing image classification and change detection [1, 2]. Differing from the ISVMs, TSVMs introduces test data in training process and considers minimizing errors produced by test dataset instead of only minimizing classification errors of training data. [1] showed substantial improvement obtained with the TSVMs over the ISVMs, especially for ill-posed classification problems caused by the limited quantity and quality of the training samples.

Gaussian process (GP) classifier which was originally developed for regression is another interesting kernel-based classification approach. By contrast to SVM classifiers, GP classifiers have not yet been investigated in the context of remote sensing. Similar to the spirit of the TSVMs that the label errors occur near the decision boundary, Lawrence and Jordan [3] proposed a Gaussian process approach, which can be viewed as a Bayesian counterpart to the TSVMs. This approach involves a novel transductive learning under a probabilistic framework to learn a GP classifier. The main difference of this approach from the standard GP is that it introduces a null category region which is equivalent to the traditional notion "margin" in TSVMs and thus the overall noise model is termed a null category noise model (NCNM). Unlike the traditional machine learning classification approaches, the NCNM maps a latent process variable *f* into three categories instead of traditionally two categories. The idea behind this approach is to use the unlabeled data to steer the decision boundary to the region of low data density. Specifically, they propose to modify the likelihood such that the model imitates the role of SVM hinge loss by penalizing the decision boundaries that lie in a high density region and favors the decision boundary with large margins.

In this paper, we propose to investigate the capabilities of NCNM for remote sensing image change detection. Regarding the change detection problem as a particular case of classification one, our aim is to detecting the changes or differences between the two multitemporal images. Given the input features $x_i \in X$, and their corresponding labels $y_i \in Y$, the goal is to learn a map $f : x_i \to y_i$ that can predict label for new coming x^* . Thus, the probability of class membership decomposes as $p(y \mid x, x^*) = \int p(y \mid f) p(f \mid x, x^*)$. The NCNM maps the hidden continuous variable *f* to three categories instead of two labels, specifically to the never used label '0' when *f* is around zero. To make use of the unlabeled data, a variable *z* is introduced to indicate whether a data is unlabeled or labeled:

$$p(z = 1 \mid y = 0) = 0 \tag{1}$$

This constraint imposes that an unlabeled data point can never take the label 0. Therefore, when the data point is unlabeled, the posterior process should be updated by $p(z \mid f)$ which is computed by $p(z = 1 \mid f) = \sum_y p(y \mid f) p(z = 1 \mid y)$.

The "effective likelihood function" for a single data point, $L(f)$, takes a piece-wise function form:

$$L(f)=\begin{cases} H(-(f+\frac{1}{2})), & for \quad y=-1, \quad z=0 \\ H(f+\frac{1}{2})-H(f-\frac{1}{2}), & for \quad z=1 \\ H(f-\frac{1}{2}), & for \quad y=+1, \quad z=0 \end{cases} \tag{2}$$

Where H is a Heaviside step function. Within the GP framework, $p(f \mid x)$ is fully specified by its mean function $m(x)$ and covariance function $k(x,x')$, expressed as:

$$p(f \mid x) \sim GP(m,k) \tag{3}$$

When $y \in \{-1,+1\}$ the effect of the likelihood will be similar to the binary classification with a GP prior. When the data is unlabeled, the effect of the likelihood will depend on the mean and variance of $p(f \mid x)$ and for all situations the effective likelihood will always force the latent function away from the null category.

Once the parameters of the process model have been learned, the prediction about the new coming x^* can be made by $p(y^*, z^* \mid x^*)$. Taking into account the constraint (1), we get:

$$p(y^* \mid x^*, z^*) \propto p(z^* \mid y^*) p(y^* \mid x^*) \tag{4}$$

Contrary to the standard Gaussian process, the NCNM allows the unlabeled data to affect the location of the decision boundary. Therefore, the predictive distribution depends not only on the labeled training data but also on the location of the test point x^*.

To thoroughly assess the NCNM performance, two different experiments are designed on the basis of several different sets of high resolution optical remote sensing images. Some change detection results are shown in *Fig. 1*. In general, the experimental results confirm the effectiveness of the proposed algorithm and also have positively indicated the NCNM classifier [3] can compete seriously with the state-of-the-art SVM^{light} classifier [4].

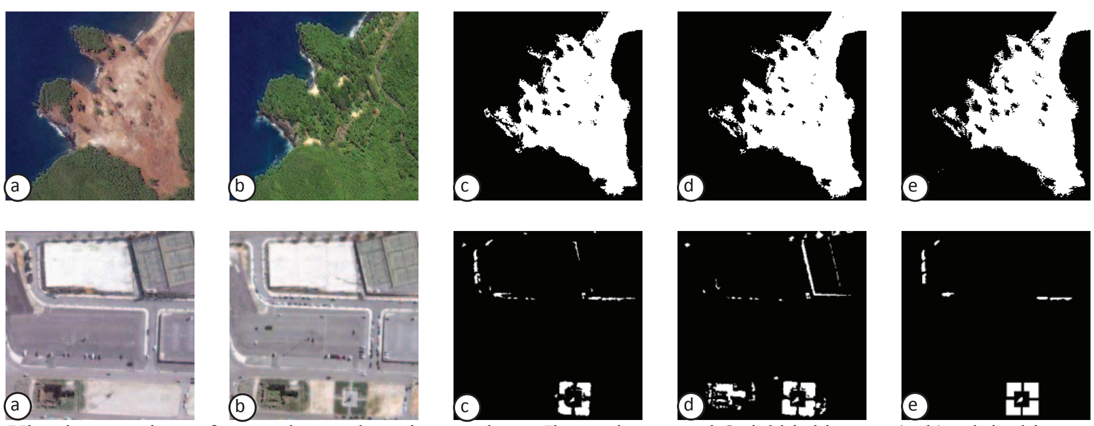

Figure 1. Visual comparison of some change detection results on Ikonos image and Quickbird image. (a, b) original images; change detection results obtained by (c) NCNM method; (d) standard TSVM classifier; (e) reference map.

REFERENCES

[1] L. Bruzzone, M. Chi, M. Marconcini. "A novel transductive SVM for semisupervised classification of remote-sensing images". IEEE Trans. Geoscience and Remote Sensing, vol. 44(11), pp. 3363-3373, Nov. 2006.

[2] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," IEEE Trans. Geoscience and Remote Sensing, vol. 42(8), pp. 1778–1790, Aug. 2004.

[3] N. D. Lawrence and M. I. Jordan. "Semi-supervised learning via Gaussian processes". L. Saul, Y. Weiss and L. Bouttou, Advances in Neural Information Processing Systems, 2005, Cambridge, MA, MIT Press.

[4] T. Joachims. "Transductive inference for text classification using support vector machines," in Proc. ICML, 1999, pp.200–209.